Cross-Assignment Discrimination in Pay: A Test Case of Major League Baseball

Őrn B. Bodvarsson[1], Kerry L. Papps [2] and John G. Sessions[3]

1 Department of Economics
Department of Management
St. Cloud State University
St. Cloud
Minnesota 56301-4498
USA
&
Institute for the Study of Labour (IZA)
Bonn, Germany

Tel: (320)308-2225
Fax: (320)255-2228
Email: obbodvarsson@stcloudstate.edu

2 Department of Economics
University of Bath
Bath BA2 7AY
UK
&
Institute for the Study of Labour (IZA)
Bonn, Germany

Tel: +44(0) 1225 384338
Fax: +44(0) 1225 383423
Email: k.l.papps@bath.ac.uk

2 Department of Economics
University of Bath
Bath BA2 7AY
UK
And
Institute for the Study of Labour (IZA)
Bonn, Germany

Tel: +44(0) 1225 384517
Fax: +44(0) 1225 383423
Email: j.g.sessions@bath.ac.uk

Abstract: The traditional Becker/Arrow style model of discrimination depicts majority and minority workers as perfectly substitutable inputs, implying that all workers have the same job assignment. The model is only appropriate for determining whether pay differences between, for example, whites and non-whites doing job assignment A are attributable to prejudice ('within-assignment discrimination'); It is inappropriate, however, for determining whether pay differences between whites in job assignment A and non-whites in job assignment B reflect discriminatory behaviour ('cross-assignment discrimination'). We test the model of such cross assignment discrimination developed by Bodvarsson and Sessions (2011) using data on Major League Baseball hitters and pitchers for four different seasons during the 1990s, a decade during which monopsony power fell. We find strong evidence of *ceteris paribus* racial pay differences between hitters and pitchers, as well as evidence that cross-assignment discrimination varies with labour market structure.

Aknowledgements: We thank William Boal, Sabien Dobbelaere, Lawrence Kahn, Wing Suen, Ted To and session participants at the 2007 SOLE and EALE meetings, the 12th IZA/CEPR European Summer Symposium in Labour Economics (ESSLE) and the 2011 RES Annual Conference, for helpful comments on earlier drafts of this paper. We also thank Bree Dority O'Callaghan and Robert Girtz for assistance with data collection and an anonymous referee and the Editor of this journal for helpful comments. The normal disclaimer applies.

Theme: Discrimination
Key Words: Wage Discrimination; Complementarity; Monopsony Power.
JEL-Code: J7

February 2014

1. Introduction

According to the U.S. Bureau of Labour Statistics, the median weekly earnings of male and female elementary and middle school teachers in 2006 were \$920 and \$824, respectively, whereas for male and female school principals and school district superintendents, median weekly earnings were \$1275 and \$1107, respectively. During the same year, the median weekly earnings of male and female registered nurses were \$1074 and \$971, respectively, whereas for male and female physicians and surgeons they were \$1847 and \$1329, respectively. Median weekly earnings of male and female cooks were \$377 and \$340, respectively, whereas for male and female restaurant waiters they were \$284 and \$348, respectively.[1]

What do the above examples have in common? First, each example involves a pair of job assignments within a firm that are distinctly *complementary*; Pilots and flight attendants are complementary labour inputs in the production of airline services, educational administrators and teachers are complements in the provision of educational services whilst physicians and nurses complement one another in the provision of health care services. Second, in each example for which data on earnings of each gender are available, there are noticeable gender pay gaps *within* job assignments – 9 per cent for school teachers, 20 per cent for principals and superintendents, nearly 10 per cent for registered nurses, 28 per cent for physicians and surgeons, nearly 30 per cent for lawyers, 10 per cent for cooks and 22.5 per cent for waitpersons (in favour of females, however). A commonly asked question would be: How much of these *intra-job* gender pay gaps are attributable to discrimination? This is the approach taken in the traditional wage discrimination model, due originally to Becker (1971) and Arrow (1973). This model is based on the fundamental assumption that majority and minority workers are perfect substitutes in production. Consequently, the traditional

[1] These numbers are taken from the Bureau of Labour Statistics website (http://www.bls.gov/cps/cpsaat39.pdf).

model is only appropriate for studying gender, racial, age, sexual orientation or other group pay differences for workers performing the same job assignment.

In this paper, we address a different and more nuanced question: To what extent is majority/minority pay *across* complementary job assignments within a firm attributable to discrimination? For example, how much of the $931 (65.6%) pay gap between male aircraft pilots and flight engineers and female transportation attendants, the $1165 pay gap between male lawyers and female legal assistants and the $876 pay gap between male physicians and surgeons and female registered nurses, attributable to discrimination? Are these gaps primarily attributable to majority/minority productivity differences or to prejudice? This is a question about *inter-job* wage discrimination and it is a far more difficult question because to answer it we need to compare majority and minority workers for which there will be both distinct productivity and labour supply differences. In the traditional (*intra-job* assignment) model of wage discrimination, details of the production function are dispensed with because there are no productivity and labour supply differences between workers. In a study of discrimination across job assignments, however, the production and labour supply functions must be given explicit consideration.

In what follows we empirically test the model of pay discrimination across job assignments developed by Bodvarsson and Sessions (2011) - hereafter BS - on an industry characterized by complementary job assignments, racial integration, variation in monopsony power across worker groups and a history of racial discrimination – U.S. Major League Baseball.

We employ a novel, two-stage regression methodology in which a standardised measure (i.e. common) measure of productivity is estimated separately for each occupation. We then incorporate this measure as a right-hand-side explanatory variable in a second-stage, all-occupation regression designed to estimate cross-assignment discrimination. Our empirical analysis finds convincing evidence of racial differences in pay across player job assignments, even after controlling for a wide array of demographic variables and position-specific productivity. Moreover, we find strong evidence of BS' theoretical prior that racial pay differentials across assignments are affected by changes in relative productivities.

The paper is set out as follows: Section 2 discusses some of the previous literature on the economics of discrimination whilst Section 4 outlines our test case of Major League Baseball. Our empirical analysis is presented in Section 5 whilst final comments are collected in Section 6.

2. Previous Literature

While most of the literature on discrimination has focused on the measurement of the majority/minority pay gap within the same job category, some researchers have suggested that the required assumption of perfect substitution between inputs may be somewhat inappropriate. Indeed, Becker alluded to this issue by sketching a brief extension to his two-factor black/white worker model to a three-factor model [see Becker (1971, pp. 59-62)]. Two of the factors are perfectly substitutable blacks and whites that belong to a group that could be termed 'Type 1 Labour.' Then, there is a third labour input, 'Type 2 Labour,' that both discriminates against blacks and is complementary to, or imperfectly substitutable, for them. Type 2 workers could, for example, be managers. In this situation, Becker showed that there would be a *ceteris paribus* black/white wage gap within the Type 1 category. Arrow (1973) elaborated on this by showing that the black/white wage gap depends upon the sensitivity of Type 2 labour's reservation wage to the fraction of the firm's labour force that is black, as well as the importance of Type 2 labour as an input (importance is measured as the size of the payments to Type 2 labour relative to Type 1 labour). Neither Becker nor Arrow tested these propositions, nor did they investigate further the implications of complementarity in production for the black/white pay differential.

Welch (1967) raised the possibility that blacks and whites working in the same firm may not be perfect substitutes because there may be differences in their educational endowments. Welch suggested that, perhaps because of long-term discrimination, blacks may have acquired less schooling and/or attended lower quality schools. He modelled educational endowments and physical labour as separate factors of production, allowing for racial differences in educational endowments and, following Becker and Arrow, white co-worker discrimination. He argued that if firms choose racially integrated labour forces then blacks and whites must be complementary inputs. The intuition is that because of whites' aversion to working with blacks, integration creates inefficiencies that will cause joint product to be less than the sum of individual black and white worker marginal products. The firm will therefore follow an apartheid employment policy unless there are sufficiently large complementarities to be exploited, i.e. if the gains from complementarity exceed the losses attributable to co-worker discrimination.[2]

More recently, Kahn (1991) sets out a model of customer discrimination in which whites and blacks are represented as different inputs in the production function. He models blacks and whites as distinct inputs because if customers are prejudiced, they will act as if the amount of black input is equal to just a fraction of the input of *otherwise identical* white workers. Similarly, Bodvarsson and Partridge (2001) present a model of a professional sports team where white and non-white athletes are imperfect substitutes due to racial differences in prior training and experience.[3]

An extensive empirical literature on wage discrimination emerged during the 1970s, all based on the original Becker-Arrow model of perfect substitution. The accumulating evidence was called into question in the early 1980s, however, as a number of studies

---

[2] Borjas (2008) also suggested that differential educational attainments may render black and white workers as imperfect substitutes: 'The two groups of workers might have different productivities because they might differ in the amount and quality of educational attainment, or because they might have been employed in different occupations and hence are entering (a) firm with different types of job training. [Borjas (2008), p. 128)].

[3] Both of these models, however, have features that limit their applicability. In Kahn's model, whites and non-whites are assigned the same job and would be perfect substitutes if customers were unprejudiced whilst Bodvarsson and Partridge impose the restriction that the cross elasticity of demand for white labour with respect to non-white labour is negative.

concluded that racial and ethnic groups were not perfectly substitutable. These studies typically applied econometric models of Translog or Generalized Leontief aggregate production functions to estimate elasticities of complementarity between groups. Grant and Hamermesh (1981), for example, found that black adults are imperfect substitutes for white men and complements to white women and youths; Borjas (1983) provided evidence which suggested that whilst black males were imperfect substitutes for white males, Hispanic males and white males were complementary; Borjas (1987) showed that black natives are imperfect substitutes for white natives; and Kahanec (2006) found that non-whites are complementary to whites.

The traditional empirical approach for testing wage discrimination is generally unsuitable where cross-assignment discrimination is concerned because it is based on a presumption that whites and non-whites are perfect substitutes. While empirical researchers have usually controlled for job assignment differences with dummy variables, that approach has severe limitations because it fails to adequately control for the structure of the underlying production function. As Hashimoto and Kochin (1980) argue, failure to account for any and all sources of productivity differences will generally exaggerate the estimated effects of discrimination.

In a recent theoretical contribution, BS extend the traditional Becker-Arrow model to ascertain how predictions regarding cross-assignment discrimination vary with the form of the production function. Using an approach similar to Kahn (1991), BS measure the extent of customer prejudice against non-white workers by a parameter, *D*.[4] Customer prejudice may be interpreted as a situation in which customers discount the marginal revenue product (MRP) of non-white workers. The lower (higher) is *D*, the more (less) intense is the prejudice and the lower (higher) is non-white MRP. Prejudice dissipates as *D* approaches 1 and reaches

[4] Note that prejudice is a necessary but not sufficient condition for *discrimination*. It is only when prejudicial thoughts are acted upon through, for example, exercising product market demand that they can result in discriminatory outcomes in the labour market. In general, taste discrimination in pay and hiring is a market outcome that results from employers acting upon their own racial preferences and/or implementing the racial preferences of customers or co-workers.

a maximum as D falls to 0. While it is traditional to think of customer discrimination as implying a price discount on the output of non-white workers, the approach above is equivalent. The parameter D reflects the idea that non-white labour is valued less when customers are prejudiced.[5] In terms of the Generalized Leontief function (GLF), for example, the impact of D is seen as follows:

$$Q = \sum_{i=1}^{k}\sum_{j=1}^{k}\gamma_{ij}[X_i^W\left(DX_j^{NW}\right)]^{\frac{1}{2}} \tag{1}$$

where $D \leq 1$, Q is output, X_i^W is the quantity of white labour input i, X_j^{NW} is the quantity of non-white labour input j, and γ_{ij} is the technology coefficient. Note there are a total of 2k inputs – two groups of workers within each job assignment (white and non-white) x k job assignments.

BO then apply Becker's (1971) *Market Discrimination Coefficient* (MDC) to the case of discrimination across job groups. The MDC measures the *ceteris paribus* racial earnings gap viz. the percentage earnings premium paid to whites. If the white and non-white wage is denoted by r^i, $i = W, NW$, then the MDC is given by:

$$MDC_{NW}^{W} = \frac{r^W(D<1)}{r^{NW}(D<1)} - \frac{r^W(D<1)}{r^{NW}(D<1)} \tag{2}$$

The first term on the right-hand side of (2) is the wage ratio when there is prejudice (i.e. when D <1) whereas the second term is the wage ratio in the absence of prejudice (i.e. when

[5] BS' approach implies that consumers can discern the racial characteristics of workers when purchasing or consuming the particular good or service in question. Such an assumption is not unrealistic and examples abound of environments in which such an approach to discounting non-white MRP is likely to hold. At professional sports events, white (non-white) fans witness non-white (white) players' contribution to athletic entertainment. If sports fans of one skin colour are prejudiced against players of another colour, this may result in lower pay to the latter group. A similar situation may arise in other entertainment services, e.g. films, theatre, popular music. More generally, there are many production situations in which consumers must interact with minority workers in order for a good or service to be dispensed, e.g. white patients interacting with non-white nurses or doctors, non-white clients interacting with white legal advisers, and white airline passengers interacting with non-white flight attendants. There will also be cases where prejudiced white consumers may not necessarily see non-white workers during the act of purchase or consumption, but mere knowledge of the racial composition of the work force may influence buying decisions. For example, white consumers may place a lower valuation on, or even refuse to purchase, food products, or appliance repair services, or the processing of important financial transactions, knowing that those goods or services were manufactured or performed by non-white workers.

D = 1). The MDC is the difference between the two ratios and measures the *ceteris paribus* racial pay gap.

Applying equation (2) to the case of cross-assignment discrimination, the *ceteris paribus* racial pay gap between whites performing job 1 and non-whites performing job 2 is:

$$MDC_{NW_2}^{W_1} = \frac{r_1^W(D<1)}{r_2^{NW}(D<1)} - \frac{r_1^W(D=1)}{r_2^{NW}(D=1)} \tag{3}$$

Similarly, the *ceteris paribus* racial pay gap between whites performing job 2 and non-whites performing job 1 is:

$$MDC_{NW_1}^{W_2} = \frac{r_2^W(D<1)}{r_1^{NW}(D<1)} - \frac{r_2^W(D=1)}{r_1^{NW}(D=1)} \tag{4}$$

BO derive the above measure of cross-assignment discrimination for four different production functions - Generalized Leontief, Quadratic, CES, and Cobb–Douglas. The Generalized Leontief provides the most general results, although closed form solutions are not possible. Closed form solutions, which are obtainable from the other three functions but only under restrictive assumptions, suggest that most predictions are generally robust across functional forms and that cross-assignment discrimination depends upon productivity and labour supply differences between the various worker groups, labour market structure, and the interaction between relative group productivity and prejudice. A uniform prediction across all four production functions is that changes in the relative productivity of one racial group induces changes in cross-assignment discrimination. For example, in all four cases, higher white (non-white) productivity raises (lowers) the amount of discrimination.[6] This is an important finding, both academically and in terms of policy. If non-whites are able to improve their skill-base, or if technological progress impacts more favorably on non-whites

[6] Whilst BS frame their theoretical model in terms of racial discrimination, it is clearly applicable to other types of labour market discrimination, for example, where workers are discriminated against on account of their age, gender, nativity status, sexual orientation, religious affiliations, or other characteristics that may be targets of employer, employee, or consumer prejudice.

relative to whites, then cross-assignment discrimination may be reduced. An increase in white productivity, however, will lead to an unintended adverse consequence by increasing discrimination against non-whites.

Table 1 following summarises BS' various comparative static results for the *ceteris paribus* white/non-white pay differential (i.e. $MDC_{NW_2}^{W_1}$) derived from the four production functions:

Table 1: BS' Comparative Static Results for Cross-Assignment Discrimination $\left(\partial MDC_{NW_2}^{W_1} / \partial Variable\right)$

Variable	*Generalized Leontief*	*Quadratic*	*CES*	*Cobb-Douglas*
Strength of Prejudice (D)	-	-	-	-
White Productivity	+	+	+	±
Non-White Productivity	-	-	-	±
White Productivity × D	-	-	-	
Non-White Productivity × D	+	+	+	
White Labour Supply			-	
Non-White Labour Supply			+	
White Reservation Wage				+
Non-White Reservation Wage				-
Employer's Monopsony Power				±
Degree Of Monopsonistic Wage Discrimination				-

BS' findings have an important general implication: Researchers must control for both productivity differences between white and non-white workers, as well as the interaction between race and productivity, when estimating the extent of cross-assignment discrimination.

2. A Test Case: Major League Baseball

In order to test empirically the BS model of cross-assignment discrimination, we searched for an appropriate test case viz. an industry where: (i) there are accurate data on salaries and productivity for individual workers across distinct job assignments and these data are available for different firms; (ii) the productivities of job assignment groups within the firm are interrelated; (iii) there is racial integration; (iv) the pay of some workers is competitively determined, whilst the pay of others is determined under conditions resembling monopsony; (v) there is potential for customer discrimination; and (vi) there have been changes in the number of employers in the industry over time.

One industry satisfying all these criteria is Major League Baseball (MLB) in the USA.[7] In MLB, each team requires two distinctly complementary types of player skill - hitting (an offensive skill) and pitching (a defensive skill) - in the production of baseball entertainment.[8] Player salaries are set under two different regimes, one competitive, the other monopsonistic. The monopsonistic regime applies to players with fewer than six years of MLB experience. These players are subject to the *reserve clause* and are constrained to negotiate their pay with only one team. The competitive regime applies to players with at least 6 years of MLB experience. They are eligible to file for *free agency* and may negotiate with any team in the league. Monopsony power effectively begins to erode, however, as early as the fourth year because then a player is eligible for *final offer arbitration*. Arbitration rights tend to relieve players of monopsonistic exploitation because arbitrators strive to award competitive salaries. Pitchers have historically been disproportionately white, whereas the pool of hitters has tended to be more racially balanced. The Major League added new teams (called 'expansion teams') since the early 1990s, leading to a reduction in each team's degree of monopsony power held over reserve clause players.

The ideal way to measure a Major League player's marginal revenue product (MRP) is by his contribution to the team's ticket, broadcasting and merchandise revenues. Because of the team production nature of baseball, however, it is difficult to empirically disentangle one player's revenue contribution from another. We thus proxy MRP by the player's years of

[7] Racial discrimination in professional sports has received considerable attention among labour economists because of the abundant statistical evidence on a player's personal attributes, compensation and productivity. Most studies in this area have focused on discrimination with respect to pay, hiring, retention and positional segregation. For an examination of the research prior to 2000, see Kahn's (2000) expository survey.

[8] Woolway (1997) and Zech (1981) argue that the Cobb-Douglas function is a particularly appropriate description of an MLB team's production situation. They both estimated Cobb-Douglas functions where the dependent variable is team winning percentage and the independent variables are player and team career statistics.

MLB experience, tenure with his current team, and various career statistics (computed on a game-by-game basis since the beginning of the player's Major League career) that proxy his ability and skills. For both hitters and pitchers we measure productivity using both individual and summary measures or performance. The individual career statistics we use to measure a hitter's productivity are: *At Bats*, *Stolen Bases*, *Bases on Balls (Walks*), *Hits*, *Sacrifice Flies*, *Hits by Pitches* and *Total Bases*. We also include the summary measure of *On Base plus Slugging* (*OPS*). We distinguish between hitters that are 'designated hitters' from those who are not. A designated hitter is a player who is chosen at the start of the game to bat in lieu of the pitcher in the line-up. We also distinguish, using dummies, between hitters that serve other types of positions. These include whether the hitter served as an infielder or a catcher. We measure a pitcher's productivity by use of the following individual career statistics: *Complete Games, Shut Outs, Saves, Home Runs Conceded, Hits by Pitches, Walks, Strike Outs, Innings Pitched and Earned Runs.* We also include the summary measure *Defence-Independent Component Earned Run Average* (*DICE*). An explanation of baseball terminology is set out in the Appendix.

3. Empirical Analysis

3.1 *Descriptive Statistics*

Tables 2 and 3 present descriptive statistics for hitters and pitchers, respectively. Our full sample comprises 1092 hitters (548 white, 367 black and 177 Hispanic) and 1204 pitchers (942 white, 127 black and 135 Hispanic). Salary, experience, performance and position data were drawn from the *Lahman Baseball Database* (see: www.baseball1.com) over four seasons - 1992, 1993, 1997 and 1998. We chose these years because the Major League expanded by two teams between 1992 and 1993 and again by two teams between 1997 and 1998. This was thus a period in which MLB teams experienced a decline in their monopsony power and as a result might be expected to have been compelled to hire more black and Hispanic players.

The salary data do not include information about contract length, bonus clauses or endorsements. Salaries for players on the Canadian teams were converted to U.S. dollars. The experience data were used to determine the player's eligibility for free agency and final offer arbitration. For the U.S. teams, metropolitan area population and per-capita income were obtained from the website of the Bureau of Economic Analysis (see: www.bea.gov). For the Canadian teams, similar data were obtained from the Statistics Canada website (see: www.statcan.ca). Per-capita income data for the Canadian cities were converted to U.S. dollars.

It would appear from Table 2 that there are no major differences between the personal and professional characteristics of white hitters, black hitters and Hispanic hitters, nor in the characteristics of the greater metropolitan area in which they play. In terms of career characteristics, however, black hitters record significantly more *At Bats*, *Stolen Bases, Bases on Balls* and *Total Basses* than either white hitters or Hispanic hitters. They are also less likely to play as an infielder or catcher, but more likely to play as an outfielder or designated hitter. Compared to Hispanic hitters, white hitters record significantly more *At Bats*, *Bases on Balls* and *Total Bases*, but significantly fewer *Stolen Bases*. They are also more likely to play as a catcher, but less likely to play as an outfielder or designated hitter.

Table 2:
Descriptive Statistic - Hitters

	All		*White*		*Black*		*Hispanic*	
Variable	*Mean*	*Std. Dev*	*Mean*	*Std. Dev*	*Mean*	*Std. Dev*	*Mean*	*Std. Dev*
Personal Characteristics								
Log Annual Salary	13.890	1.13	13.865	1.10	13.938	1.13	13.866	1.22
Age	30.304	3.70	30.596	3.49	30.488	3.95	29.023	3.55
White	0.502	0.50	-	-	-	-	-	-
Black	0.336	0.48	-	-	-	-	-	-
Hispanic	0.162	0.37	-	-	-	-	-	-
Professional Characteristics								
MLB Experience	7.061	3.89	7.062	3.87	7.223	4.07	6.723	3.55
MLB Experience-Squared	64.957	69.31	64.785	70.06	68.684	74.23	57.763	54.59
Tenure with Current Club	2.672	3.00	3.062	3.38	2.305	2.62	2.226	2.24
Free Agent	0.600	0.49	0.598	0.49	0.605	0.49	0.599	0.49
Eligible for Final Offer Arbitration	0.296	0.46	0.304	0.46	0.294	0.46	0.271	0.45
American League	0.514	0.50	0.521	0.50	0.469	0.50	0.588	0.49

National League	0.486	0.50	0.479	0.50	0.057	0.23	0.124	0.33
Canadian Team	0.073	0.26	0.067	0.25	7.223	4.07	6.723	3.55
Performance								
At Bats	2506.414	2001.58	2419.738	1940.51	2699.202	2198.95	2375.525	1720.23
Stolen Bases	69.746	112.52	44.800	72.35	111.055	157.89	61.480	69.63
Bases on Balls (Walks)	254.275	247.74	253.131	233.32	285.349	293.87	193.39	161.14
Hits	693.882	589.16	658.164	559.63	749.278	639.78	689.605	562.82
Sacrifice Flies	22.946	21.52	22.891	21.16	23.569	22.31	21.825	21.04
Hits by Pitch	17.663	17.97	18.389	18.57	17.095	16.83	16.593	18.36
Total Bases	1060.200	913.52	1016.772	880.39	1162.845	1013.19	982.073	771.85
On Base Plus Slugging (OPS)	0.742	0.09	0.740	0.08	0.755	0.09	0.723	0.11
Infielder	0.459	0.50	0.556	0.50	0.281	0.45	0.531	0.50
Outfielder	0.383	0.49	0.217	0.41	0.657	0.48	0.333	0.47
Catcher	0.116	0.32	0.189	0.39	0.016	0.13	0.096	0.30
Designated Hitter	0.059	0.24	0.046	0.21	0.079	0.27	0.056	0.23
Greater Metro Area Characteristics								
Percentage White	80.507	6.89	80.938	6.77	80.683	6.72	78.808	7.39
Percentage Black	13.273	6.58	12.959	6.60	13.676	6.62	13.409	6.44
Percentage Hispanic	10.621	10.65	10.719	10.80	10.331	10.58	10.918	10.36
Average Annual Income ($)	25562.990	3789.65	25508.570	3757.99	25551.300	3731.59	25756.00	4016.17
Population[1]	5514009	4657988	5313189	4509095	5513759	4729589	6137413	4927354
Year Dummies								
1992	0.250	0.43	0.255	0.44	0.243	0.43	0.249	0.43
1993	0.235	0.42	0.248	0.44	0.237	0.43	0.192	0.40
1997	0.260	0.44	0.248	0.43	0.270	0.44	0.277	0.45
1998	0.255	0.44	0.250	0.43	0.251	0.43	0.282	0.45
Sample Size	1092		548		367		177	

Notes:
1. Population denotes the greater metro area population;
2. Source: All variables except Greater Metro Area Characteristics (GMAC) extracted from the Lahman Baseball Database (Version 5.0, Release Date: Dec. 15, 2002). GMAC derived from the Statistical Abstract 1997-1999, the BEA, CA1-3, and from Statistical Canada..

Table 3:
Descriptive Statistics - Pitchers

Variable	All		White		Black		Hispanic	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Personal Characteristics								
Log Annual Salary	13.409	1.19	13.451	1.20	13.238	1.16	13.276	1.18
Age	29.815	4.09	30.190	4.02	29.016	4.00	27.948	4.03
White	0.782	0.41	-	-	-	-	-	-
Black	0.105	0.31	-	-	-	-	-	-
Hispanic	0.162	0.37	-	-	-	-	-	-
Professional Characteristics								
MLB Experience	5.988	4.20	6.158	4.20	5.772	4.49	5.000	3.75
MLB Experience-Squared	53.468	76.64	55.562	78.38	53.331	75.31	38.985	63.34
Tenure with Current Club	1.924	2.07	1.935	2.10	1.843	1.97	1.926	1.99
Free Agent	0.467	0.50	0.482	0.50	0.441	0.50	0.385	0.49
Eligible for Final Offer Arbitration	0.306	0.46	0.314	0.46	0.236	0.43	0.319	0.47
American League	0.513	0.50	0.518	0.50	0.543	0.50	0.452	0.50
National League	0.487	0.50	0.475	0.50	0.528	0.50	0.556	0.50
Canadian Team	0.069	0.25	0.063	0.24	0.055	0.23	0.126	0.33
Performance								
Complete Games	10.15	22.24	10.981	23.33	6.433	14.87	7.844	19.65
Shutouts	2.875	6.08	3.065	6.32	1.984	4.74	2.385	5.35
Saves	19.488	51.87	20.941	52.93	19.362	62.60	9.474	26.16
Home Runs Conceded	56.517	62.57	58.842	64.46	50.409	52.94	46.044	56.11
Hits by Pitches								
Walks	225.779	249.73	231.782	257.66	224.095	217.58	185.474	217.41
Strikeouts	436.641	514.13	450.726	530.21	436.047	490.18	338.919	402.35
Innings Pitched	627.59	702.43	655.160	720.78	558.969	620.14	499.785	627.21
Earned Run Average	4.025	0.96	3.995	0.94	4.175	1.11	4.094	0.97
DICE	4.172	0.71	4.159	0.71	4.250	0.67	4.189	0.77
Starter	0.442	0.50	0.441	0.50	0.402	0.49	0.489	0.50
Greater Metro Area Characteristics								
Percentage White	80.714	6.84	80.695	6.91	80.335	6.56	81.201	6.59
Percentage Black	13.038	6.46	12.946	6.49	14.026	6.46	12.750	6.19
Percentage Hispanic	10.975	10.77	10.899	10.61	10.909	10.40	11.573	12.20
Average Annual Income ($)	25488.2	3939.85	25491.51	3895.30	25852.23	3898.44	25122.19	4271.98
Population[1]	5551948	4683875	5481401	4631793	6035905	4915887	5588930	4829139
Year Dummies								
1992	0.221	0.42	0.236	0.42	0.189	0.39	0.148	0.36
1993	0.239	0.43	0.248	0.43	0.244	0.43	0.170	0.38
1997	0.264	0.44	0.256	0.44	0.276	0.45	0.311	0.46
1998	0.276	0.45	0.260	0.44	0.291	0.46	0.370	0.48
Sample Size	1204		942		127		135	

Notes:
1. Population denotes the greater metro area population;
2. Source: All variables except Greater Metro Area Characteristics (GMAC) extracted from the Lahman Baseball Database (Version 5.0, Release Date: Dec. 15, 2002). GMAC derived from the Statistical Abstract 1997-1999, the BEA, CA1-3, and from Statistical Canada
3. DICE = Defence-Independent Component Earned Run Average

In Table 3, the domination of white pitchers is immediately apparent. White pitchers are on average older than both black and (especially) Hispanic pitchers. They also enjoy higher average earnings. In terms of career characteristics, white pitchers record significantly higher *Wins*, *Losses*, *Games Started*, *Complete Games*, *Shutouts*, *Saves*, *Homeruns*, *Walks*, *Strikeouts* and *Innings Pitched* than either blacks or Hispanic pitchers, with Hispanic pitchers recording generally lower figures than black pitchers.

3.2 *Empirical Methodology*

Wage discrimination occurs when individuals who are identical in terms of their productive characteristics are paid differently on account of their non-productive characteristics. Any empirical analysis of discrimination thus requires some control of productivity - it would not be surprising, and nor would it suggest discrimination, if more productive individuals were paid more than less productive individuals. In the traditional literature such control is usually straightforward since the individuals under scrutiny are performing the same job. In our model, however, it is problematic. Our concern is whether there is discrimination *across* job assignments, that is, where individuals with *different* non-productive characteristics are performing *different* jobs - do male airline pilots earn more than female flight attendants because of their occupation or because of their gender? This is a difficult issue to address empirically because we need to control for the productivity of both the pilot and the flight attendant or, more generally, we need to control for assignment-specific productivity. Clearly some measures of productivity will be common across job assignments - for example, education, job-tenure, and labour market experience. By definition, however, some measures of productivity will be unique to particular job assignments and it is controlling for these that is the real challenge.

One possible solution is to adopt a two-stage generated regressor approach.[9] Assume that wages reflect productivity as follows: To ascertain the level of discrimination across player positions, we need to control for position-specific productivity. In one sense this is straightforward because some measures of off-field productivity (MLB experience and tenure with current team, for example) are common across pitchers and hitters. On-field measures of productivity, however, vary across hitters and pitchers; e.g. runs for hitters and strike-outs for pitchers. Given our objective of ascertaining the extent of racial discrimination *across* job assignments, we need a standardized productivity measure. We thus adopt the following two-stage approach. We first assume that wages reflect productivity as follows:

$$\ln w^{ij} = \mathbf{X}_0^j \mathbf{B}_0^{ij} + \mathbf{X}_1 \mathbf{B}_1^{ij} \tag{5}$$

$\ln w^{ij}$ denotes the log wage of a member of group $i = 1,2,...,I$ employed in job assignment $j = 1,2,...,J$, $\mathbf{X}_0^j$ is a vector of 'assignment-specific' productivity measures, $\mathbf{X}_1$ is a vector of 'common' (i.e. cross assignment) productivity measures (e.g. education, tenure), and the **B**'s denote parameter vectors. Our aim is to derive an estimating equation of the form:

$$\ln w^{ij} = \mathbf{X}_0 \mathbf{B}_0^{ij} + \mathbf{X}_1 \mathbf{B}_1^{ij} \tag{6}$$

where $\mathbf{X}_0$ denotes some standardised (imputed) measure of assignment-specific productivity. To this end, we estimate the following 'first-stage' group-assignment regressions:

$$\ln w^{j} = \mathbf{X}_0^j \mathbf{A}_0^{j} \tag{7}$$

That is, we estimate separate wage regressions for individual within each job assignment, including as explanatory variables only the various assignment-specific productivity

[9] See Pagan (1984), Gauger (1989) and Gawande (1996) for discussions of the inference issues regarding estimated regressor models.

measures. Thus, we estimate separate wage regressions for hitters and pitchers on only their respective position-specific variables in both individual (I) and summary (S) form:

Hitters: (I) *At Bats*; *Stolen Bases*; *Bases on Balls (Walks)*; *Hits, Sacrifice Flies, Hits by Pitches, Total Basses*, *Infielder*; *Outfielder*; *Catcher*; *Designated Hitter*.
(S) *On Base plus Slugging* (*OBS*); *Infielder*; *Outfielder*; *Catcher*; *Designated Hitter.*

Pitchers: (I) *Complete Games*; *Shut Outs*; *Saves*; *Home Runs Conceded*; *Hit by Pitches*; *Walks*, *Strike Outs*; *Innings Pitched*; *Earned Runs*; *Starter*.
(S) *Defence-Independent Component Earned Run Average (DICE); Starter*.

We then use the predicted values from these regressions, $\hat{w} = \left(\hat{w}^{ij}; \forall i, j\right)$, as a standardized measure of assignment-specific productivity in second-stage regressions of the form:

$$\ln w^{ij} = \hat{w}\mathbf{B}_0^{ij} + \mathbf{X}_1\mathbf{B}_1^{ij} \tag{8}$$

3.3 *Cross-Assignment Regression Analysis*

Our estimates of equation (7) - the first stage regression - for hitters and pitchers respectively are set out in Tables 4 and 5 following:

Table 4:
First-Stage Regression – Hitters

	Individual Measures (1)		*Summary Measures* (2)	
Variable	*Coef.*	*Std. Error*	*Coef.*	*Std. Error*
At Bats	-0.000	0.000	-	-
Stolen Bases	0.000	0.000	-	-
Walks	0.000	0.000	-	-
Hits	-0.000	0.000	-	-
Sacrifice Flies	0.003	0.004	-	-
Hit By Pitches	0.007***	0.002	-	-
Total Bases	0.001***	0.000	-	-
Infielder	0.142	0.238	0.041	0.269
Outfielder	0.087	0.235	-0.178	0.266
Catcher	0.265	0.250	0.116	0.282
Designated Hitter	-0.265	0.199	0.179	0.223
On Base Plus Slugging (OPS)	-	-	6.377***	0.352
Constant	12.915***	0.239	9.183***	0.379
R-Squared	0.427		0.246	
Number of Observations	1092		1092	

Note: *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 5:

First-Stage Regression - Pitchers

	(1) Individual Measures		(2) Summary Measures	
	Coef.	*Std. Error*	*Coef.*	*Std. Error*
Complete Games	-0.063***	0.009	-	-
Shut Outs	0.113***	0.033	-	-
Saves	0.002***	0.001	-	-
Home Runs Conceded	-0.008**	0.003	-	-
Hit By Pitches	0.001	0.001	-	-
Walks	0.000	0.001	-	-
Strike Outs	0.003***	0.001	-	-
Innings Pitched	-0.001	0.001	-	-
Earned Runs	0.000	0.001	-	-
Starter	0.300***	0.08	2.359***	0.369
Starter × Complete Games	0.025***	0.010		
Starter × Shut Outs	-0.054	0.034		
Starter × Saves	-0.001	0.003		
Starter × Home Runs Conceded	0.008**	0.004		
Starter × Hit By Pitches	0.003*	0.001		
Starter × Walks	-0.000	0.001		
Starter × Strike Outs	-0.001	0.001		
Starter × Innings Pitched	-0.002	0.002		
Starter × Earned Runs	-0.002	0.001		
DICE[1]	-	-	-0.561***	0.051
Starter × DICE[1]	-	-	-0.387***	0.087
Constant	12.348***	0.051	15.423***	0.215
R-Squared		0.547		0.272
Number of Observations		1204		1204

Notes: 1. *DICE ~ Defence-Independent Component ERA;* 2. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

In the first specification of Table 4, the individual hitter performance measures are used as regressors. A player's total number of bases has a significant positive effect on his salary. Controlling for this, players who obtain a lot of their bases through singles (i.e. have a lot of hits or at bats) have lower salaries, however this effect is not significant. Being hit by a pitch is also a significant determinant of salary. In the second specification of the table, the individual measures are replaced with *OPS*, which is found to have a highly significant positive effect on salary.

In the first specification of Table 5, pitchers' individual performance measures are included in the salary regression. Strike outs have a highly significant positive effect on salary. Shutouts, saves and conceding few home runs are also important. Having a lot of complete games is seen to lower relief pitchers' salaries, but have no effect on starters' salaries. Holding everything else equal, starters are seen to earn more than relief pitchers. When *DICE* is used in place of the individual performance measures in specification 2, it is found to have a significant negative effect on salary, as predicted, and this is found to be stronger for starters than for relief pitchers.

Table 6 provides an analysis of how salary and productivity (as estimated in Tables 4 and Table 5) varies across position and race/ethnicity groups:

Table 6:
Mean Salary and Productivity for Race / Position Groups

	Whites	*Blacks*	*Hispanics*
Log Salary			
Pitchers	13.451	13.238*†††	13.276†††
Hitters	13.865	13.938†††	13.866†††
Log Salary minus Productivity			
Pitchers	0.021	-0.205***††	0.046
Hitters	-0.024	0.007	0.059

Notes: *, ** and *** denote significant difference from whites in the same position at the 10%, 5% and 1% level, respectively. †, †† and ††† denote significant difference from whites in the opposite position at the 10%, 5% and 1% level, respectively.

In the first panel of Table 6 we compare the mean (log) salary across race / position groups. In all cases hitters earn more than pitchers but the gap is particularly large for blacks and Hispanics. Black pitchers earn significantly less than white pitchers, black and Hispanic pitchers earn significantly less than white hitters whilst black and Hispanic hitters earn significantly more than white pitchers. The key question is whether these cross-assignment gaps are due to systematic differences across races in pitching and hitting ability or whether they reflect discrimination with equally productive players being remunerated differently depending upon race. In the second panel of Table 6 we compare mean (log) salary less estimated productivity across race / position groups. It is apparent that most of the cross-assignment differences are now insignificant, suggesting that the differences in salary may be largely explained by differences in performance. However, black pitchers appear to be underpaid relative to their productivity compared to both white pitchers and white hitters.

The values in the second panel implicitly assume that whites, blacks and Hispanics are all rewarded at the same rate for given increases in productivity. The theory of BS, however, suggests that this may not be case. In addition, the values do not control for components of productivity that affect both pitchers and hitters. Therefore, we set out our estimate of equation (8) (i.e. the second stage regression) in Table 7 and Table 8 following.

In Table 7 we report four second-stage specifications with log salary being regressed on imputed productivity from Table 4 and Table 5, dummy variables for race and position, interactions between productivity, race and position, and a set of controls for professional characteristic that affect the salaries of both pitchers and hitter. Whites are the reference category in all specifications. In specifications (1)-(3), imputed productivity is derived from individual productivity measures [i.e. specification (1) in Table 4 and Table 5] whilst specification (4) is derived from the relevant summary measure employed in specification (2) of Table 4 (i.e. *OPS*) and Table 5 (i.e. *DICE*).[10]

Despite the tendency to focus on 'hitters versus pitchers' baseball remains a team sport and we allow for possible complementarities in production by introducing team fixed effects in specifications (2)-(4) of Table 7 to control for both differences in teams' abilities to pay and team spill-overs.[11] We also control for the possibility that discrimination reflects a team's willingness to only select players of races that are common in the local area by including in specification (3) the percentages of each race category living in the greater metropolitan area in which the team is located.

The results in Table 7 suggest strong evidence of both cross- and within-assignment discrimination in MLB. Our estimated coefficients are robust across the four specifications and suggest that, at the mean of productivity, white hitters earn 0.14% more than black

[10] Using OPS in the first stage imposes the implicit constraint that slugging average and on-base percentage have the same effect on earnings, which is unlikely to be realistic. However, when slugging average and on-base percentage were used in place of OPS, the second stage results were almost identical.

[11] For example, Gould and Winter (2010) show that a worker's effort has a positive effect on the effort of coworkers if they are complements in production and a negative effect if they are substitutes. In terms of MLB specifically, they find that a hitter's productivity is positively related to the productivity of the other hitters on his team but negatively related to his team's pitching quality.

pitchers; 2.13% less than Hispanic pitchers; 0.69% less than black hitters; 0.14% less than Hispanic hitters; and 1.55% less than white pitchers.

The coefficients on the triple interactions (i.e. *position* × *race* × *productivity*) capture how much a given increase in productivity is rewarded in the labour market within a specific position-race group relative to its effect on white hitters. It would appear that up to a certain level of productivity, our results support BS' predictions. Consider, for example, black pitchers versus white hitters: We know from Table 6 that black pitchers are underpaid relative to white hitters. In Table 7, we see a positive and significant positive coefficient on black*pitcher and a negative and significant coefficient on black*pitcher*productivity. This suggests that at low productivity levels, black pitchers are overpaid compared to whites and that, in this region, an increase in black pitcher productivity will decrease the MDC. Beyond a certain performance level, however, the earnings of black pitchers fall below those of white hitters. In this region, blacks pitchers are relatively underpaid, and raising their productivity will increase the MDC. Comparing Hispanic pitchers to white hitters, we find the opposite relationship with Hispanic pitcher being relatively underpaid (overpaid) at low (high) levels of productivity. The two cases are illustrated graphically in Figure 1 with black and Hispanic pitcher pay equalling that of white hitters at productivity levels of 13.57 and 12.12 respectively:

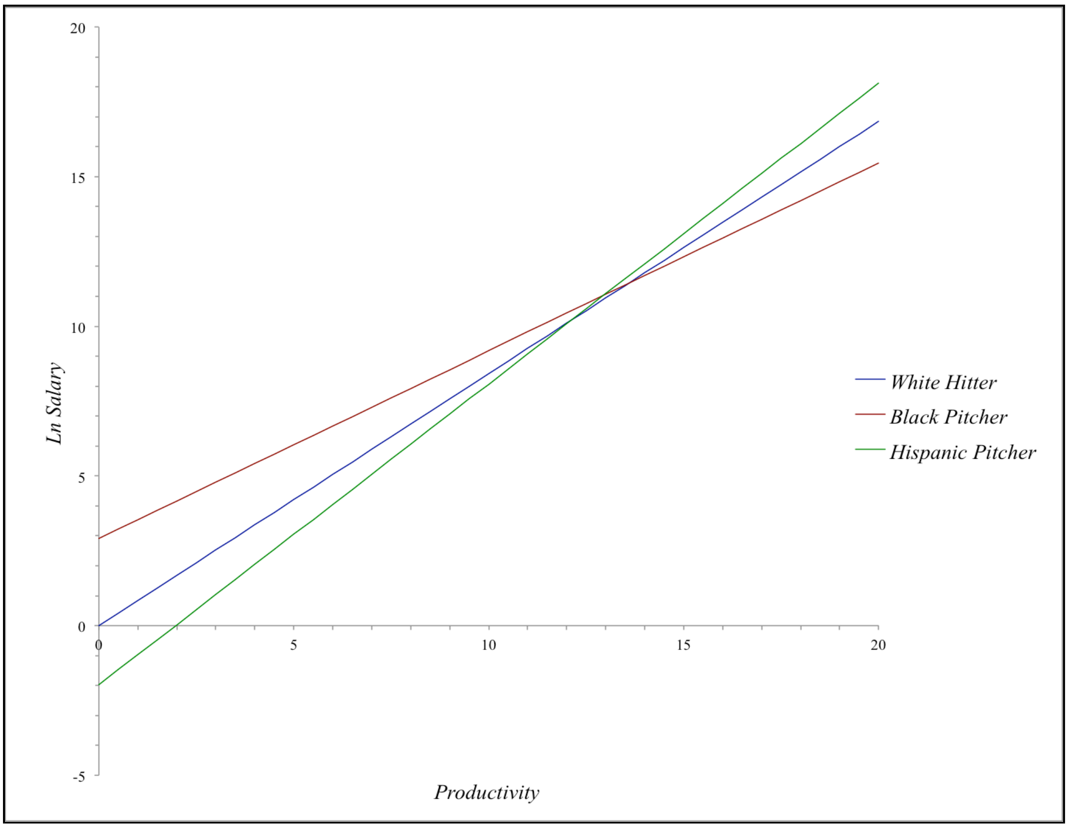

Figure 1: Cross Assignement Wage Discrimination and Productivity

Four additional regressions are set out in Table 8. Specification (1) groups U.S.-born Hispanics with whites (using Hispanic classification based on players' facial appearances). Specification (2) uses players' performance measures from the previous season, rather than career performance to date, in the first-stage regressions since it might be expected that these have a greater effect on salary than performance several seasons before. Finally, to explore how discrimination varies with the degree of monopsony power, specifications (3) and (4) split the sample into free agents and non-free agents respectively.

It would appear from Table 8 that our results are robust to treating U.S. born Hispanics as whites and to estimating productivity on the previous season's performance rather than on career performance to date. Finally, we find evidence that discrimination is higher amongst non-free agents, a result that would appear to contradict BS' prior that discrimination may decline with the degree of monopsony power in the market.

Table 7: Second-Stage Regression Results

Dependent Variable: Log Annual Salary

	(1)		(2)		(3)		(4)	
Variable	*Coefficient*	*Std. Error*	*Coefficient*	*Std. Error*	*Coefficient*	*Std. Error*	*Coefficient*	*Std. Error*
(Imputed) Productivity								
Individual Measures	0.842***	0.148	0.845***	0.146	0.848***	0.144	-	-
Summary Measure	-	-	-	-	-	-	0.606***	0.032
Interactions								
Pitcher	0.181	0.911	0.274	0.852	0.309	0.828	-0.866	0.434
Pitcher × Black	2.723***	0.220	2.438***	0.309	2.519***	0.324	2.842***	0.186
Pitcher × Hispanic	-2.157***	0.181	-2.143***	0.070	-2.156***	0.065	0.655	0.354
Pitcher × Productivity	-0.000	0.065	-0.007	0.060	-0.010	0.058	0.080**	0.031
Pitcher × Black × Productivity	-0.214***	0.017	-0.192***	0.023	-0.198***	0.024	-0.217***	0.014
Pitcher × Hispanic × Productivity	0.163***	0.013	0.162***	0.005	0.163***	0.006	-0.053*	0.025
Hitter × Black	-0.348***	0.412	-0.308***	0.352	-0.244***	0.354	-1.347***	0.104
Hitter × Hispanic	-4.061**	1.127	-4.154**	1.063	-4.165***	1.002	-1.749***	0.344
Hitter × Black × Productivity	0.031	0.031	0.028	0.027	0.023	0.027	0.105***	0.008
Hitter × Hispanic × Productivity	0.299**	0.083	0.305**	0.078	0.306***	0.074	0.135***	0.025
Professional Characteristics								
Age	-0.034*	0.013	-0.030*	0.014	-0.029	0.015	-0.033*	0.015
Experience	0.152**	0.051	0.140**	0.051	0.137**	0.052	0.304***	0.043
Experience Squared	-0.010**	0.003	-0.009**	0.003	-0.009**	0.003	-0.012***	0.002
Tenure	0.068**	0.021	0.069**	0.019	0.068**	0.019	0.078***	0.011
Free Agent	0.947***	0.132	0.970***	0.132	0.980***	0.134	0.746***	0.110
Eligible for Final Offer Arbitration	0.523***	0.071	0.534***	0.067	0.537***	0.071	0.391***	0.058
American League	-0.040	0.034	-0.077	0.071	0.040	0.060	-0.128***	0.026
Canadian Team	-0.100	0.092	-	-	-	-	-	-
Year Dummies								
1993	0.068**	0.020	0.064***	0.014	0.082**	0.021	0.073**	0.026
1997	0.114	0.082	0.034	0.076	-0.043	0.039	0.099	0.112
1998	0.208*	0.081	0.108	0.060	0.072**	0.025	0.190	0.118
Greater Metro Area Characteristics								
Percentage White	-	-	-	-	0.041	0.023	-	-
Percentage Black	-	-	-	-	0.218***	0.035	-	-
Percentage Hispanic	-	-	-	-	0.083**	0.021	-	-
Average Annual Income	0.000*	0.000	0.000*	0.000	0.000	0.000	0.000	0.000
Metropolitan Area Population	0.000	0.000	-0.000	0.000	-0.000*	0.000	0.000	0.000
Constant	1.543	1.883	1.246	2.278	-4.602	3.299	3.621**	0.986
Team Fixed Effects	No		Yes		Yes		Yes	
R-Squared	0.692		0.700		0.702		0.673	
Number of Observations	2296		2296		2296		2296	

Notes: 1. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively;
2. *White Hitter* is the reference category.

Table 8: Additional Second-Stage Regression Results

Dependent Variable: Log Annual Salary

Variable	(1)		(2)		(3)		(4)	
	US Born Hispanics = Whites		*Previous Season*		*Free Agents*		*Non-Free Agents*	
	Coefficient	*Std. Error*	*Coefficient*	*Std. Error*	*Coefficient*	*Std. Error*	*Coefficient*	*Std. Error*
(Imputed) Productivity								
Individual Measures	0.798***	0.138	0.816***	0.034	0.771***	0.128	2.420***	0.068
Pitcher	0.535	0.853	0.778	0.343	0.355	1.079	16.019***	0.391
Pitcher × Black	2.128***	0.326	0.987***	0.098	1.003	0.412	5.692***	0.186
Pitcher × Hispanic	-2.674***	0.507	-0.332*	0.163	-5.022***	0.735	1.639**	0.512
Pitcher × Productivity	-0.026	0.060	-0.050	0.025	-0.014	0.076	-1.171***	0.029
Pitcher × Black × Productivity	-0.161***	0.025	-0.078***	0.007	-0.094*	0.029	-0.440***	0.014
Pitcher × Hispanic × Productivity	0.196***	0.036	0.027	0.012	0.356***	0.050	-0.120*	0.040
Hitter × Black	-0.385	0.261	0.221*	0.072	0.264	0.433	-0.347	0.742
Hitter × Hispanic	-3.906*	1.304	-2.049***	0.487	-3.439*	1.327	-3.049*	0.767
Hitter × Hispanic × Productivity	0.284*	0.095	0.152***	0.035	0.248*	0.095	0.234***	0.057
Hitter × Black × Productivity	0.034	0.020	-0.013	0.006	-0.012	0.032	0.038	0.055
R-Squared	0.683		0.787		0.479		0.744	
Number of Observations	2083		2296		1217		1079	

Notes: 1. All specifications include the regressors from specification (3) of Table 7;
2. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively;
3. White Hitter is the reference category;
4. Specification (1) groups US-born Hispanics with whites, using Hispanic classification based on players' facial appearances. Specification (2) incorporates (individual) performance measures from the previous season in the first-stage regressions. Specifications (3)-(4) split the sample into free agents and non-free agents respectively.

3.4. Decomposition Analysis

In this section, we attempt to identify cross-assignment discrimination using another empirical approach. The fact that players of a particular race in a particular position enjoy a wage differential over players of another race in another position could be a reflection of the former group's greater endowment of 'earning characteristics'. White pitchers may, for example, be more productive or have more experience on average than non-white (i.e. black or Hispanic) hitters. Alternatively, white pitchers may be better rewarded for the characteristics they do possess, suggesting some form of positive (negative) discrimination from employers towards white pitchers (non-white hitters). To address this issue we perform a Blinder-Oaxaca *decomposition* to separate the earnings differential into an 'endowment component', to account for differences in endowments between individuals, and a 'price component', which is usually associated with discrimination.[12]

Recalling equation (7), we write the earnings function of players of race *j* in position *i* as:

$$\ln w^{ij} = \mathbf{X}^{ij}\mathbf{B}^{ij} + \varepsilon^{ij} \tag{9}$$

where $i = (W, NW)$ and $j = (H, P)$ denote white and non-white and pitchers and hitters respectively, and where $NW = (B, H)$ denotes black and Hispanic respectively. $\mathbf{X}^{ij} = \left(\mathbf{X}_0^{ij}, \mathbf{X}_1^{ij}\right)$ denotes our vectors of position-specific and common productivity characteristics, $\mathbf{B}^{ij} = \left(\mathbf{B}_0^{ij}, \mathbf{B}_1^{ij}\right)$ the corresponding coefficient vectors to be estimated, and ε^{ij} some well-behaved error term. Thus, the earnings functions of white pitchers, non-white pitchers, white hitters and non-white hitters may be denoted:

[12] This method of decomposition, initially proposed by Oaxaca (1973) and Blinder (1973), and later generalized by Oaxaca and Ransom (1994), has been applied extensively to discrimination on the basis of gender, race, caste and religion.

$$\ln w^{WP} = \mathbf{X}^{WP}\mathbf{B}^{WP} + \varepsilon^{WP} \tag{10}$$

$$\ln w^{NWP} = \mathbf{X}^{NWP}\mathbf{B}^{NWP} + \varepsilon^{NWP} \tag{11}$$

$$\ln w^{WH} = \mathbf{X}^{WH}\mathbf{B}^{WH} + \varepsilon^{WH} \tag{12}$$

$$\ln w^{NWH} = \mathbf{X}^{NWH}\mathbf{B}^{NWH} + \varepsilon^{NWH} \tag{13}$$

The Blinder-Oaxaca decomposition divides wage differentials into a part that is 'explained' by group differences in productivity and a residual part that cannot be accounted for by such differences in wage determinants. This latter 'unexplained' component is often used as a measure for discrimination. For example, the predicted average white pitcher/non-white hitter (WP-NWH) differential may be represented as:

$$\begin{aligned}
&\Delta \ln w^{WP-NWH} = \ln w^{WP} - \ln w^{NWH} = \bar{\mathbf{X}}^{WP}\hat{\mathbf{B}}^{WP} - \bar{\mathbf{X}}^{NWH}\hat{\mathbf{B}}^{NWH} \\
&\Rightarrow \\
&\Delta \ln w^{WP-NWH} = \hat{\mathbf{B}}^{NWH}\left(\bar{\mathbf{X}}^{WP} - \bar{\mathbf{X}}^{NWH}\right) + \bar{\mathbf{X}}^{WP}\left(\hat{\mathbf{B}}^{WP} - \hat{\mathbf{B}}^{NWH}\right)
\end{aligned} \tag{14}$$

The first term, $\hat{\mathbf{B}}^{NWH}\left(\bar{\mathbf{X}}^{WP} - \bar{\mathbf{X}}^{NWH}\right)$, represents differences in endowments between members of the two groups whilst the second term, $\bar{\mathbf{X}}^{WP}\left(\hat{\mathbf{B}}^{WP} - \hat{\mathbf{B}}^{NWH}\right)$, represents differences in coefficients and thus, by extension, rewards. Note that if the overall differential is negative (i.e. $\Delta \ln w^{WP-NWH} < 0$) but the second term is positive [i.e. $\bar{\mathbf{X}}^{WP}\left(\hat{\mathbf{B}}^{WP} - \hat{\mathbf{B}}^{NWH}\right) > 0$], then it would suggest that non-white hitters are discriminated against despite earning, on average, more than white hitters - i.e. non-white hitters would do even better with the earnings generating function of white pitchers than with their own.

Table 9 reports the above decomposition analysis on the salary gaps between each cross-assignment pair. Focussing on the mean decomposition results first, it is apparent from the first and second panels of Table 8 that our regression model implies a positive salary premium for black and Hispanic hitters over white pitchers *ceteris paribus*. The decomposition suggests that this premium would be even greater in the absence of discrimination, with discrimination *against* black and Hispanic hitters alleviating the potential differential by almost 50 per cent and 100 per cent respectively.

The third and fourth panels focus on the white hitter / black pitcher and white hitter / Hispanic pitcher decomposition. Both decompositions imply a positive mean salary premium at for white hitters with discrimination playing a relative minor role in the white hitter / black pitcher differential, discrimination *against* white hitters reducing the potential white hitter premium by just under 5 per cent. It would appear that discrimination plays a much more significant role in the white hitter / Hispanic pitcher differential with discrimination *against* white hitters reducing the potential differential by just over 60 per cent.

The results set out in Table 9 suggest that the coefficients effect generally works to reduce the minority group's salary, even where that group has higher average pay. However, the quantile decomposition suggests that that at the bottom of the salary distribution over 100% of cross-assignment differences between race groups is explained by differences in observable productivity, with differences in rewards working to reduce the gap. At the top end of the salary distribution, the coefficients effect is mostly insignificant, a finding that is consistent with a situation in which salaries are compressed at the bottom end of the distribution [Papps (2013)].

Table 9

Decomposition Analysis

Race / Position Group Pair	*Mean*		*10th Percentile*		*50th Percentile*		*90th Percentile*	
	Coefficient	*%*	*Coefficient*	*%*	*Coefficient*	*%*	*Coefficient*	*%*
White Pitcher – Black Hitter								
Endowment Effect	-0.727***	*149.28*	-0.618***	*143.39*	-0.581***	*104.87*	-0.334***	*112.08*
Coefficients Effect	0.240***	*-49.28*	0.187***	*-43.39*	0.027	*-4.87*	0.036	*-12.08*
Total Effect	-0.487***	*100.00*	-0.431***	*100.00*	-0.554***	*100.00*	-0.298***	*100.00*
White Pitcher – Hispanic Hitter								
Endowment Effect	-0.828***	*199.52*	-0.569***	*241.10*	-0.543***	*93.78*	-0.194**	*64.03*
Coefficients Effect	0.413***	*-99.52*	0.333***	*-141.10*	-0.036 **	*6.22*	-0.109	*35.97*
Total Effect	-0.415***	*100.00*	-0.236***	*100.00*	-0.579***	*100.00*	-0.303***	*100.00*
White Hitter – Black Pitcher								
Endowment Effect	0.599***	*95.53*	1.122***	*218.71*	0.946***	*111.82*	-0.086	*-21.13*
Coefficients Effect	0.028	*4.47*	-0.609***	*-118.71*	-0.100	*118.20*	0.493**	*121.13*
Total Effect	0.627***	*100.00*	0.513***	*100.00*	0.846***	*100.00*	0.407***	*100.00*
White Hitter – Hispanic Pitcher								
Endowment Effect	0.947***	*160.78*	0.949***	*212.30*	1.058***	*122.60*	0.437***	*191.67*
Coefficients Effect	-0.358***	*-60.78*	-0.502***	*-112.30*	-0.195***	*-22.60*	-0.209***	*91/67*
Total Effect	0.589***	*100.00*	0.447***	*100.00*	0.863***	*100.00*	0.228	*100.00*

Notes: 1. The mean results refer to Blinder-Oaxaca decompositions of the differences in log salary between the two groups and use the coefficients of the second group in each case; 2. The percentile results refer to decompositions of the differences between the quantile function of log salary for the two groups, as described in Chernozhukov *et al*. (2009); In all cases, the regressors include age, estimated productivity, experience, experience squared, tenure, free agent, salary arbitration eligibility, American League, Canadian team, average annual income, metropolitan area population, year dummies and a constant; 3. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Our decomposition results are also consistent with our regression results from the previous section. That is, for the white hitter / black pitcher decomposition, whilst the coefficients effect explains part of the overall wage gap (cf. the first column of Table 9 with the second panel of Table 6), at the 10^{th} (90^{th}) percentile the coefficients effect reduces (widens) inequality, as illustrated in Figure 1. As regards the white hitter / Hispanic pitcher decomposition, the coefficients effect implies that Hispanic hitters earn more than they should, given their productivity, at both the 10^{th} and 90^{th} percentiles, a finding that is also consistent with Figure 1 assuming that Hispanic hitters are always in the upper region.

6. Final Comments

In this study we address a widely neglected problem in the literature on taste discrimination in pay: ascertaining the extent to which racial or gender differences in pay across job assignments are attributable to prejudice. Virtually every wage discrimination study has focused on discrimination within the same job assignment, thus treating whites and non-whites (or males and females) as perfect substitutes. In a recent contribution, BS investigate the case of discrimination *across* job assignments, where assignments are viewed as distinct inputs. BS' theoretical findings underscore the importance of carefully considering the production function when there are productivity differences between majority and minority workers. An important finding from their theoretical analysis is that the magnitude of, for example, differences in white relative to non-white productivity influences the amount of discrimination. Furthermore, when whites and non-whites are interrelated in production, race and productivity will interact. This is an important implication, for it means that *whenever* white and non-white workers have productivity differences, the researcher should include productivity x race interactions in any empirical specification.

We test BS' model using data from Major League Baseball, an industry characterized by complementary job assignments, a history of racial integration and discrimination, and a dual labour market structure. We find convincing evidence of racial differences in pay across player job assignments, even after controlling for a wide array of demographic variables and position-specific productivity. Moreover, we find strong evidence of BS' theoretical prior that racial pay differentials across assignments are affected by changes in relative productivities.

References

Arrow, K. J. (1973). 'The Theory of Discrimination.' In O.A. Ashenfelter and A. Rees (Eds.), *Discrimination in Labour Markets*. Princeton NJ: Princeton University Press, pp. 3-33.

Becker, G. S. (1971). *The Economics of Discrimination, Second Edition.* Chicago: University of Chicago Press.

Blinder, A.S. (1973). 'Wage Discrimination: Reduced Form and Structural Estimates.' *Journal of Human Resources*, 2, pp. 8-22.

Bodvarsson, Ö. B. and J. G. Sessions. (2011). 'The Measurement of Pay Discrimination Between Job Assignments.' *Labour Economics*. 18 pp. 645-657.

Bodvarsson, Ö. B. and M. D. Partridge. (2001). 'A Supply and Demand Model of Employer, Customer and Co-worker Discrimination.' *Labour Economics*, 8, June, pp. 389-416.

Borjas, G. J. (1983). 'The Substitutability of Black, Hispanic, and White Labour.' *Economic Inquiry*, 21, January, pp. 93-106.

Borjas, G. J. (1987). 'Immigrants, Minorities, and Labour Market Competition.' *Industrial and Labour Relations Review*, 40 (3), April, pp. 382-92.

Borjas, G. J. (2008). *Labour Economics*, 4th edition. Dubuque, IA: McGraw-Hill Irwin.

Chernozhukov. V., I. Fernandez-Val and B. Melly (2009). 'Inference on Counterfactual Distributions.' CeMMAP working papers CWP09/09, *Centre for Microdata Methods and Practice, Institute for Fiscal Studies*.

Cotton, J. (1988). 'On the Decomposition of Wage Differentials.' *Review of Economics and Statistics*, 70, pp. 236-243.

Gauger, J. (1989). The Generated Regressor Correction: Impacts upon Inferences in Hypothesis Testing.' *Journal of Macroeconomics*, 11(3), pp.383-395.

Gawande, K. (1997). Generated Regressors in Linear and Nonlinear Models.' *Economics Letters*, 54, pp119-126.

Gould, E. and E. Winter (2009). 'Interactions Between Workers and the Technology of Production: Evidence from Professional Baseball.' *Review of Economics and Statistics*, 91(1), pp. 188-200.

Grant, J.H. and D.S. Hamermesh (1981). 'Labour Market Competition Among Youths, White Women and Others.' *The Review of Economics and Statistics*, 63(3), August, pp. 354-60.

Hashimoto, M. and L. Kochin (1980). 'A Bias in the Statistical Effects of Discrimination.' *Economic Inquiry*, 8, pp. 478-86.

Holmes, P. (2011). 'New Evidence of Salary Discrimination in Major League Baseball.' *Labour Economics*, 18(3), pp. 320-331.

Kahanec, M. (2006). 'The Substitutability of Labour of Selected Ethnic Groups in the U.S. Labour Market.' *IZA Discussion Paper No. 1945*, January (www.iza.org).

Kahn, L. M. (1991). 'Customer Discrimination and Affirmative Action.' *Economic Inquiry*, 29, July, pp. 555-571.

Kahn, L. M. (2000). 'A Level Playing Field? Sports and Discrimination.' In William S. Kern (ed.), *The Economics of Sports*, W.E. Upjohn Institute, Kalamazoo, MI.

Neumark, D. (1988). 'Employers' Discriminatory Behavior and the Estimation of Wage Discrimination.' *The Journal of Human Resources*, 23, pp.279–295.

Oaxaca, R. L. (1973). 'Male-Female Wage Differentials in Urban Labour Markets.' *International Economic Review*, 14, pp. 337-356.

Oaxaca, R. L. and M. R. Ransom. (1994) 'On Discrimination and the Decomposition of Wage Differentials.' *Journal of Econometrics*, 61(1), pp. 5-21.

Pagan, A. (1984). 'Econometric Issues in the Analysis of Regressions with Generated Regressors.' *International Economic Review*, 25(1), pp. 221-247

Papps, K. L. (2013). 'Productivity over the Life Cycle: Evidence from Professional Baseball.' SOLE 18th Annual Meeting, Boston, 3-4 May, 2013.

Reimers, C. W. (1983). 'Labour Market Discrimination against Hispanic and Black Men.' *The Review of Economics and Statistics*, 65, pp. 570–579.

Welch, F. (1967). 'Labour-Market Discrimination: An Interpretation of Income Differences in the Rural South.' *Journal of Political Economy*, 75(3), June, pp. 225-40.

Woolway, M.D. (1997). 'Using an Empirically Estimated Production Function for Major League Baseball to Examine Worker Disincentives Associated with Multi-Year Contracts.' *The American Economist*, 41(2), 77-84.

Zech, C.E. (1981). 'An Empirical Estimation of a Production Function: The Case of Major League Baseball.' *The American Economist*, 25(2), 19-23.

Appendix: Baseball Terminology

General

1. An *infielder* is a defensive player who plays on the *infield*, the dirt portion of a baseball diamond between first and third bases. The specific infielder positions are *first baseman*, *second baseman*, *shortstop* (which is between second and third bases) and *third baseman*. In contrast, an *outfielder* plays farthest from the batter and his primary role is to catch long fly balls. Outfielder positions include *left fielder*, *centre fielder* and *right fielder*. The *catcher* crouches behind home plate and receives the ball from the *pitcher*. Because the *catcher* can see the whole field, he is best positioned to lead and direct his fellow players in play. He typically calls the pitches by means of hand signals, hence requires awareness of both the pitcher's mechanics and the strengths and weaknesses of the batter. A *designated hitter* is a player permitted to bat in place of the pitcher in the American League.

Hitters

2. A player has an *at bat* every time he comes to bat, except in certain circumstances, e.g. if he is awarded first base due to interference or obstruction or the inning ends while he is still at bat.
3. A hitter is assigned a *stolen base* (also called a *steal*) when he reaches an extra base on a hit from another player. For example, suppose that hitter A is at first base when hitter B hits the ball. Hitter B reaches first base (he would be assigned a *single*), but hitter A reaches third base. Hitter A would be assigned a *stolen base* because he reached an extra base.
4. A *base on balls* (also called a *walk*) is assigned when the batter receives four pitches each of which the umpire determines is a *ball*. A *ball* is any pitch at which the batter does not swing and is out of the *strike zone* (which means it would not qualify to be a *strike*). When the hitter is assigned a *base on balls*, he is entitled to walk to first base.
5. A batter scores a *hit* when he safely reaches first base after hitting the ball into fair territory, without the benefit of an error or a fielder's choice.
6. A *sacrifice fly* is a batted ball that satisfies four criteria: (i) there are fewer than two outs when the ball is hit; (ii) the ball is hit to the outfield (fair or foul) or to infield foul territory; (iii) the batter is put out because an outfielder (or an infielder running in the outfield or foul territory) catches the ball on the fly (or alternatively if the batter would have been out if not for an error or if the outfielder drops the ball and another runner is put out); and (iv) a runner who is already on base scores on the play. It is called a 'sacrifice' fly because the batter presumably intends to cause a teammate to score a run while sacrificing his own ability to do so. A sacrifice fly is not counted as a turn at bat for the batter, though the batter is credited with a run batted in.
7. *Hits by pitches* (*HBP*) records the number of times that a batter or his equipment (other than his bat) is hit in some part of his body by a pitch from the pitcher. A hit batsman is awarded first base, provided that (in the plate umpire's judgment) he made an honest effort to avoid the pitch.
8. *Total bases* are the number of bases a player has gained through hitting. It is the sum of his *hits* weighted by 1 for a *single*, 2 for a *double* (if he gets to second base as a result of his hit), 3 for a *triple* (if he gets to third base) and 4 for a *home run*. Only bases attained from hits count toward this total. After a player collects a hit, whether it be a single, double, triple or home run, the total bases stat can be applied. Whether or not this player advances further during the inning, by stealing a base or advancing off another players hit, does not increase his/her total base number. Thus, $TB = 1B + 2(2B) + 3(3B) + 4(HR)$

9. *On-base plus slugging* (*OPS*) is a hitter's *slugging average plus his on-base percentage* (*OBP*). *Slugging average* reflects hitting power and defined as the ratio of *total bases* to *at bats*. *On-base percentage* (*OBP*) is a measure of how often a batter reaches base for any reason other than a fielding error, fielder's choice, dropped / uncaught third strike, fielder's obstruction, or catcher's interference (the latter two are ignored as either times-on-base (TOB) or plate appearances in calculating *OBP*).
10. *OBP* is calculated as: $OBP = \left(H + BB + HBP\right)/\left(AB + BB + HBP + SF\right)$.
11. *OBS* is calculated as: $OBS = \left[AB\left(H + BB + HBP\right) + TB\left(AB + BB + HBP + SF\right)\right]/\left[AB\left(AB + BB + HBP + SF\right)\right]$.

Pitchers

12. A *pitcher* is assigned a *win* or a *loss* depending on whether he was the *pitcher of record* when the decisive run was scored. One is the *pitcher of record* if one is the *pitcher* at the point when the player who scores the decisive run is allowed to reach a base.
13. *Complete games is* the number of games where the pitcher was the only pitcher for his team.
14. A pitcher earns a *save* if he enters a game led by the pitcher's team, finishes the game without surrendering the lead, is not the winning pitcher, and either (a) the lead was three runs or fewer when the pitcher entered the game; (b) the potential tying run was on base, at bat, or on deck; or (c) the pitcher pitched three or more innings.
15. Number of *home runs conceded* (*HR*) which is assumed to be negatively related to salary, is the number of pitches that were hit by batters which were scored as a *home run*.
16. A *pitcher* is assigned a *walk*, which is assumed to be negatively related to salary, if he allows a batter to reach base after pitching him four balls. He is assigned a *strikeout* (*K*) if he pitches three *strikes* (pitched balls counted against the batter, typically swung at and missed or fouled off) in a row.
17. An *inning* is one of nine periods in a MLB game in which each team has a turn at bat; *innings pitched* is the number of such periods when the pitcher was working.
18. *Earned run average* is negatively correlated with the pitcher's ability to prevent the opposing team from scoring. It equals the number of times the pitcher allows a batter to score a *Run* (where the batter scores a point by advancing around the bases and reaching home plate safely) multiplied by nine, divided by the number of *Innings Pitched*.
19. *Defense-Independent Component ERA* (*DICE*) was created in 2001 as a variation on *Component ERA* (*CERA*). *CERA* was created to create a more accurate way of evaluating pitchers than *ERA*. Whereas *ERA* is significantly affected by luck (such as whether the component hits are allowed consecutively), *CERA* eliminates this factor and assigns a weight to each of the recorded 'components' of a pitcher's performance. For *CERA*, these are singles, doubles, triples, home runs, walks and hit batters. *DICE* is an improvement on *CERA* that removes the contribution of the pitcher's defense and instead estimates a pitcher's *ERA* from the components of his pitching record that do not involve defense. These are *home runs*, *walks*, *hits by pitch* and *strikeouts*. DICE is thus calculated as: $DICE = 3 + \left[13HR + 3\left(BB + HBP\right) - 2K\right]/IP$